Emerging Molecular Technologies for Identifying the Risk of Second Cancers

Perspective on Wu et al., p. 617

Susan T. Mayne and Stephen B. Gruber

Cancer is becoming an increasingly survivable disease, and the number of cancer survivors is increasing each year. Second or higher-order cancers now account for about 16% of incident cancers in the Surveillance, Epidemiology, and End Results database (1). An important current research need is to understand which patients are at greatest risk for second cancers and, thus, the greatest potential beneficiaries of available or experimental targeted surveillance and chemoprevention strategies.

Research over the past decade has emphasized the development of cancer risk models based on host characteristics (e.g., age, sex, reproductive factors), exogenous exposures (e.g., human papillomavirus, environmental tobacco smoke), lifestyle factors (e.g., smoking, alcohol), and (for second cancers) tumor characteristics (e.g., disease site, stage). These models have moved the field of cancer risk prediction forward with regard to both first and second cancers. In this issue of the journal, the field moves forward again with the report of Wu et al. (2), which examines a large number of single nucleotide polymorphisms in relation to the risk of second primary tumors/local recurrence in patients who have been curatively treated for early-stage squamous cell carcinomas of the oral cavity, pharynx, and larynx, the first large-scale evaluation of this type in head and neck cancer patients. The results are striking. The authors offer data supporting the hypothesis that risk prediction significantly and substantially improves following the incorporation of genetic data (multiple risk loci) into prediction models and that second cancers are partially attributable to polygenic variation, as has been clearly shown for first primary head and neck cancers (3).

Squamous cell carcinoma of the head and neck has been a model disease site for research in second primary prevention for several reasons. Patients with early-stage disease frequently are treated curatively, surviving their disease to be at risk for second cancers. The annual risk for second primary cancers is 5% among stage I/II patients (4). Second cancers and local recurrences are a major cause of morbidity and mortality in this patient population (4). Various chemopreventive approaches to reduce this risk of second cancers have met with disappointing results. Retinoids, β-carotene, α-tocopherol, and N-acetylcysteine have been evaluated (5) in large-scale cancer prevention trials. Although chemoprevention trials in head and neck cancer have not been efficacious, these same trials are now contributing importantly to the understanding of the etiology of second head and neck cancer as investigators, such as Wu et al. (2), begin to mine the data from prospectively followed and carefully characterized patient populations.

The sample of patients in the Wu et al. analysis had histologically confirmed stage I or II head and neck squamous cell carcinoma and participated in the Retinoid Head and Neck Second Primary Trial, a randomized, placebo-controlled chemoprevention trial of daily low-dose 13-cis-retinoic acid in preventing second primary tumors over several years (6). The trial was well designed and conducted but disproved the hypothesis that this particular retinoid could prevent second primary tumors overall—the hazard ratio for second primary tumors in the 13-cis-retinoic acid versus placebo group was 1.06 (95% confidence interval, 0.83-1.35). Wu et al. (2) used a nested case-control approach in comparing single-nucleotide polymorphisms (SNP) in DNA from 150 patients who developed second primary tumors and local recurrences with SNPs from 300 patients who remained free of second primary tumors and local recurrences throughout the follow-up period. This ability to convert a null chemoprevention trial into an informative cohort study is an often unrecognized value of well-conducted chemoprevention trials—an important value considering how many trials have had disappointing findings in their primary efficacy analyses.

Wu et al. (2) did not use the increasingly popular genome-wide association (GWAS) approach to identify target variants of interest, using instead a candidate gene approach, where they examined a large number of genes thought to be related to carcinogenesis. They selected haplotype-tagging SNPs, SNPs in coding and regulatory regions, and functional SNPs based on allele frequency and other factors, leading to the inclusion of 8370 chromosomal and mitochondrial SNPs in nearly 1,000 cancer-related genes for analysis. Less than half of the cases with second primary cancer or recurrence had DNA available for analysis, so the investigators wisely chose to focus on candidate genes because of the limited sample size. They then explored how these SNP’s predict future disease status using a variety of statistical approaches.

First, they examined the associations between SNPs and outcome (adjusted for multiple comparisons), identifying 6 chromosomal and 7 mitochondrial SNPs of interest. They tested each SNP under recessive, dominant, and additive models. It could be argued that the authors’ choice to evaluate each

Authors’ Affiliations: 1Yale School of Public Health and Medicine, New Haven, Connecticut and 2University of Michigan Medical School and School of Public Health, Ann Arbor, Michigan

Received 5/11/09; accepted 5/29/09.

Grant support: Grant No. P50 CA97248 from the NIH through the University of Michigan Head and Neck Specialized Program of Research Excellence.

Note: Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerprevention.aacrjournals.org/).

Requests for reprints: Susan T. Mayne, Yale School of Public Health, 60 College Street, P.O. Box 208034, New Haven, CT 06520-8034. Phone: 203-785-6274; Fax: 203-785-6980; E-mail: susan.mayne@yale.edu.

© 2009 American Association for Cancer Research.

tol.1158/1940-6207.CAPR-09-0101

www.aacrjournals.org

model increased the chances of false discovery because many investigators choose to rely on a single genetic model that provides a parsimonious statistical test when the underlying mode of inheritance is unknown (7, 8). However, the authors went further to diminish the effect of multiple comparisons.

For example, they examined the consistency of these associations via a bootstrap resampling procedure (100 times), finding that 12 chromosomal SNPs and 3 mitochondrial SNPs had a bootstrap P value <0.01 80% of the time. Two of the three mitochondrial SNPs were in high linkage disequilibrium with the third and thus were excluded from further analyses, leaving 13 SNPs having the most consistent associations with a poor outcome. The authors then examined the effect of carrying more versus fewer unfavorable genotypes by summing the number of unfavorable genotypes in the 13 selected SNPs. Carrying more of these unfavorable genotypes versus fewer was strongly predictive of outcome: Persons with eight or more unfavorable genotypes were at 27-fold increased risk of second primary cancer or recurrence compared with those with four or fewer unfavorable genotypes. Comparing receiver-operator characteristic curves that included the genotype data with receiver-operator characteristic curves that only included known clinical and smoking variables revealed that adding genotype data significantly improved the sensitivity and specificity of outcome prediction [as measured by area under the curve (AUC)], with the AUC increasing from 0.64 (clinical plus smoking variables) to 0.81 (clinical, smoking, and genetic variables).

Another sophisticated statistical approach of Wu et al. in examining multilevel interactions of the various genotypes was a survival-tree analysis (using only chromosomal SNPs, apparently). The tree and final risk nodes (Supplementary Fig. S1) are based on quite small sample sizes, and thus the order of selected SNPs and the SNP profiles that drive each node must be interpreted with caution. It becomes immediately evident on examination of the tree, however, that second cancer risk is determined by the complex interplay of numerous genetic polymorphisms and that the risk conferred by a particular risk polymorphism can be readily offset by other genetic polymorphisms, suggesting that simple models of genetic risk (for low-frequency allelic variants) likely are not forthcoming.

Replication of findings as important and provocative as the present results of Wu et al. (2) is essential, albeit difficult to accomplish in this circumstance. The International Head and Neck Cancer Epidemiology Consortium is an international consortium to support research in head and neck cancer epidemiology. However, this consortium emphasizes first primary cancers. There are few, if any, existing resources in the world that have the same high-quality clinical data from a cohort of head and neck cancer patients, and none that we are aware of with longer follow-up. New technologies, however, may provide a window into even more data from studies like the Retinoid Head and Neck Second Primary Trial (6). Now that a list of candidate genes has been proposed, it may be possible to extract DNA from archived paraffin-embedded tissues from both cases and controls, all of whom are first primary cancer patients, to independently validate the Wu et al. findings. The quality and quantity of DNA from paraffin are usually insufficient to permit high-throughput genotyping of nearly 9,000 SNPs. Yet, the DNA derived from paraffin samples is sufficient to genotype the 13 SNPs identified by Wu et al. (2) in very large, independent series of cases. Paraffin-derived DNA should permit these investigators to conduct a validation study among the 204 second primary and recurrent cases with retrievable blocks who could not be evaluated in the present study because of a lack of lymphocytes, as well as permitting validation studies by other research teams with large collections of archived tissue accompanied by good follow-up data.

The finding that incorporating data on selected genetic polymorphisms with more standard (lifestyle, host) factors can substantially improve the prediction of cancer risk has not been consistently observed in the broader cancer literature. For example, Zheng et al. (9) found that adding information on five SNPs into risk models for prostate cancer increased the AUC under the receiver-operator characteristic curve from 0.624 to only 0.633 (a statistically significant but relatively minor improvement), versus a model based on age, geographic region, and family history of prostate cancer. Similarly, Gail (10) found that adding seven SNPs identified in recent GWAS studies to the National Cancer Institute Breast Cancer Risk Assessment Tool (which includes age at first live birth, age at menarche, number of first-degree relatives with breast cancer, and number of previous benign breast biopsy examinations) improved the AUC for the National Cancer Institute risk assessment model from 0.607 to 0.632, also a relatively small change. Spitz et al. (11) used a slightly different approach, developing a risk model for lung cancer based on traditional risk factor data (environmental tobacco smoke, family history of cancer, dust exposure, prior respiratory disease, and smoking history variables) and then expanding the model with two markers of DNA repair capacity (mutagen sensitivity and host-cell reactivation; ref. 12). Incorporating the biomarker data significantly increased the AUCs, although the change was relatively modest (from 0.67 to 0.70 for former smokers and from 0.68 to 0.73 for current smokers). A critical difference between these studies and that of Wu et al. (2) is that these other studies included family history data. That is, incorporation of family history data in risk models may lessen the value of adding genotypic/phenotypic biomarker data to them, a hypothesis that can be readily tested in populations with available family history and genotypic data.

A critical question that was not answered by the analysis of Wu et al. (2) is whether they could identify genetic subgroups that may benefit most from the chemopreventive intervention (13-cis-retinoic acid) tested in their source trial. The authors commented on a preliminary finding that one variant in the MKI67 (cell cycle) gene was associated with retinoid efficacy; that is, it was both prognostic and predictive of treatment response, suggesting the possibility of personalized chemoprevention. This preliminary finding, as with all subgroup findings, must be interpreted with caution. However, there are other completed retinoid-based chemoprevention trials in head and neck cancer (13, 14) that could attempt to replicate this pharmacogenomic finding, assuming they have available DNA and with the caveat that they evaluated three different retinoids.

This report of Wu et al. has many implications. First, it highlights the value of collecting biospecimens along with lifestyle

http://www.inhance.iarc.fr
and demographic data in chemoprevention trials, whether primary prevention trials or trials aimed at second cancer prevention (see ref. 1). Second, it suggests that individualized data on allelic variation in common genes can significantly improve risk prediction, especially when family history is not already accounted for in the risk model. Third, it opens a window for other head and neck cancer studies to replicate its prognostic associations involving a very limited panel of SNPs. It also may be possible for other completed studies to try to replicate the pharmacogenomic finding involving one SNP, although these studies may lack statistical power to accomplish this. Last, it highlights the substantial progress that can be made by a capable team of multidisciplinary investigators using new technologies to ask timely questions. Molecular epidemiology has had its challenges (15), but Wu et al. (2) have shown that this field continues to provide new insights, including a clearer understanding of disease etiology along with enhanced possibilities for individualized risk prediction.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

References
Emerging Molecular Technologies for Identifying the Risk of Second Cancers

Susan T. Mayne and Stephen B. Gruber


Updated version
Access the most recent version of this article at:
http://cancerpreventionresearch.aacrjournals.org/content/2/7/605

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2009/07/13/2.7.605.DC1

Cited articles
This article cites 12 articles, 7 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/2/7/605.full.html#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
/content/2/7/605.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.